

Approximating capacitated k -median with $(1 + \epsilon)k$ open facilities

Shi Li*

Abstract

In the capacitated k -median (CKM) problem, we are given a set F of facilities, each facility $i \in F$ with a capacity u_i , a set C of clients, a metric d over $F \cup C$ and an integer k . The goal is to open k facilities in F and connect the clients C to the open facilities such that each facility i is connected by at most u_i clients, so as to minimize the total connection cost.

In this paper, we give the first constant approximation for CKM, that only violates the cardinality constraint by a factor of $1 + \epsilon$. This generalizes the result of [Li15], which only works for the uniform capacitated case. Moreover, the approximation ratio we obtain is $O(\frac{1}{\epsilon^2} \log \frac{1}{\epsilon})$, which is an exponential improvement over the ratio of $\exp(O(1/\epsilon^2))$ in [Li15]. The natural LP relaxation for the problem, which almost all previous algorithms for CKM are based on, has unbounded integrality gap even if $(2 - \epsilon)k$ facilities can be opened. We introduce a novel configuration LP for the problem, that overcomes this integrality gap.

*TTIC, shili@ttic.edu

1 Introduction

In the capacitated k -median (CKM) problem, we are given a set F of facilities, each facility i with a capacity $u_i \in \mathbb{Z}_{>0}$, a set C of clients, a metric d over $F \cup C$ and an upper bound k on the number of facilities we can open. The goal is to find a set $S \subseteq F$ of at most k open facilities and a connection assignment $\sigma : C \rightarrow S$ of clients to open facilities such that $|\sigma^{-1}(i)| \leq u_i$ for every facility $i \in S$, so as to minimize the connection cost $\sum_{j \in C} d(j, \sigma(j))$.

A special case of the problem is the classic k -median (KM) problem, in which all facilities $i \in F$ have $u_i = \infty$. There has been a steady stream of papers on approximating KM [4, 10, 11, 18, 19, 24]. The current best approximation ratio for the problem is $2.611 + \epsilon$ due to Byrka et al. [8]. On the negative side, it is NP-hard to approximate the problem within a factor of $1 + 2/e - \epsilon \approx 1.736$ [18].

Very little is known about the CKM problem. Most previous algorithms [1, 7, 11, 14, 21] are based on the natural LP relaxation for the problem, which has unbounded integrality gap even if all capacities are the same, and one is allowed to violate the cardinality constraint (the constraint that at most k facilities can be open) *or* the capacity constraints by a factor of $2 - \epsilon$. As a result, these algorithms can only give pseudo-approximation algorithms for CKM. Indeed, for algorithms based on the natural LP relaxation, near-optimal cardinality violation and capacity violation constants have been obtained by Gijswijt and Li [1] and [21], respectively. [1] gave a $(7 + \epsilon)$ -approximation for CKM that opens $2k$ facilities. [21] gave an $O(1/\epsilon)$ -approximation with $2 + \epsilon$ factor violation on the capacity constraints. The first algorithm that breaks the factor of 2 barrier on the cardinality violation factor is given by Li [22]. [22] gave an $\exp(O(1/\epsilon^2))$ -approximation algorithm with $(1 + \epsilon)$ -violation in the number of open facilities for the uniform CKM, the special case of CKM when all facilities i have the same capacity $u_i = u$. His algorithm is based on a stronger LP called “rectangle LP”.

There are two slightly different versions of CKM. The version as we described is the hard CKM problem, where each facility can be opened once; while in the soft CKM problem, multiple copies of each facility can be open. Hard CKM is more general as one can convert a soft CKM instance to a hard CKM instance by making k copies of each facility. The result of [14] is for soft CKM while the other mentioned results are for hard CKM. In particular, [22] showed that when the capacities are the same, hard CKM and soft CKM are equivalent, up to a constant loss in the approximation ratio.

Our contributions We generalize the result of [22] to the non-uniform CKM problem. Moreover, we improve the approximation ratio exponentially from $\exp(O(1/\epsilon^2))$ in [22] to $O(\frac{1}{\epsilon^2} \log \frac{1}{\epsilon})$. To be more precise, we give an $\left(1 + \epsilon, O\left(\frac{1}{\epsilon^2} \log \frac{1}{\epsilon}\right)\right)$ -approximation algorithm for soft CKM with running time $n^{O(1/\epsilon)}$, where an (α, β) -approximation algorithm is an algorithm which outputs a solution with at most αk open facilities whose cost is at most β times the optimum cost with k open facilities.

Theorem 1.1. *Given a soft capacitated CKM instance, and a parameter $\epsilon > 0$, we can find in $n^{O(1/\epsilon)}$ time a solution to the instance with at most $(1 + \epsilon)k$ open facilities whose cost is at most $O(\frac{1}{\epsilon^2} \log \frac{1}{\epsilon})$ times the cost of optimum solution with k open facilities.*

As most previous results are based on the basic LP relaxation for the problem, which has unbounded integrality gap even if we can open $(2 - \epsilon)k$ facilities, our result gives the first constant approximation algorithm for CKM which only violates the cardinality constraint by $1 + \epsilon$ factor. Compared to the result of [22], our algorithm works for general capacities and gives an exponentially better approximation ratio. On the downside, the running time of our algorithm is $n^{O(1/\epsilon)}$, instead of $n^{O(1)}$. It only works for soft-capacitated k -median: we may open 2 copies of the same facility due to some technicality.

Overview Our algorithm is based on a novel configuration LP; we highlight the key ideas here. The following bad situation is a barrier for the basic LP relaxation. There is an isolated group of facilities

and clients in the metric. The fractional solution uses $y \geq 1$ fractional open facilities to serve all clients in the group. However, the integral solution needs to use at least $\lceil y \rceil$ facilities. $\lceil y \rceil / y$ can be as large as $2 - \epsilon$. To bound the number of open facilities by $(1 + \epsilon)k$, we need to handle the bad case when $y \leq \ell_1$ for some large enough $\ell_1 = \Theta(1/\epsilon)$. For the uniform capacitated case, this was handled by the “rectangle LP” in [22]. However, the LP heavily depends on the uniformity of the capacities and generalizing it to the non-uniform case seems very hard.

Instead, we use a more complicated configuration LP. Let B be the set of facilities in the group. If we know that the optimum solution opens at most ℓ_1 facilities in B , we can afford to use $n^{O(\ell_1)}$ configuration variables and natural constraints to characterize the convex hull of all valid integral solutions w.r.t B . For example, for any $S \subseteq B$ of size at most ℓ , we can use z_S^B to indicate the event that the set of open facilities in B is exactly S .

But we do not know if at most ℓ_1 facilities are open in B . To overcome this issue, we also have a variable z_\perp^B to indicate the event that number of open facilities in B is more than ℓ_1 . When we condition on $z_\perp^B = 0$, we obtain a convex hull of integral solutions w.r.t B . If we condition on $z_\perp^B = 1$, we only get a fractional solution w.r.t B to the basic LP. However, this is good enough, as the bad situation corresponds to $z_\perp^B = 0$.

We have constraints for every subset $B \subseteq F$ in our configuration LP. As a result, the LP is hard to solve. We use a standard trick that has been used in many previous papers (e.g. [3, 9, 22]). Given a fractional solution, our rounding algorithm either finds a good integral solution, or finds a violated constraint. This allows us to combine our rounding algorithm with the ellipsoid method.

Our rounding algorithm uses the framework of [22]. We create a clustering of facilities, each cluster with a center that is called a representative. Then we partition the set of representatives into groups, and bound the number of open facilities and the connection cost via a group-by-group analysis. The improved approximation ratio comes from a novel partitioning algorithm, where we partition the representatives based on the minimum spanning tree over these representatives.

Organization In Section 2, we introduce the basic and the configuration LP relaxations for CKM. Then in Section 3, we give an $(O(1), O(1))$ -approximation for CKM based on the basic LP. The result is not new and the constants are worse than those in [1]. However, the algorithm serves as a starting line for our $\left(1 + \epsilon, O\left(\frac{1}{\epsilon^2} \log \frac{1}{\epsilon}\right)\right)$ -approximation algorithm. Then, in Section 4, we give the rounding algorithm based on the configuration LP. We end this paper with some open problems in Section 5. All omitted proofs can be found in Appendix A.

2 The basic LP and the configuration LP

In this section, we give our configuration LP for CKM. We start with the following basic LP relaxation:

$$\begin{aligned} \min \quad & \sum_{i \in F, j \in C} d(i, j) x_{i,j} \quad \text{s.t.} \quad & \text{(Basic LP)} \\ \sum_{i \in F} y_i & \leq k; & (1) \quad & \sum_{j \in C} x_{i,j} \leq u_i y_i, \quad \forall i \in F; & (4) \\ \sum_{i \in F} x_{i,j} & = 1, \quad \forall j \in C; & (2) \quad & 0 \leq x_{i,j}, y_i \leq 1, \quad \forall i \in F, j \in C. & (5) \\ x_{i,j} & \leq y_i, \quad \forall i \in F, j \in C; & (3) \end{aligned}$$

In the above LP, y_i indicates whether facility i is open or not, and $x_{i,j}$ indicates whether the client j is connected to facility i . Constraint (1) restricts us to open at most k facilities, Constraint (2) requires every client to be connected to a facility, Constraint (3) says that a client can only be connected to an open facility and Constraint (4) is the capacity constraint. In the integer programming, we require $x_{i,j}, y_i \in \{0, 1\}$ for every $i \in F, j \in C$. In the LP relaxation, we relax the constraint to $x_{i,j}, y_i \in [0, 1]$.

In the soft-capacitated case, y_i may be any non-negative integer. We can make k copies of each facility i and assume the instance is hard-capacitated. Thus we assume $y_i \in \{0, 1\}$ in a valid integral solution.

The above LP has unbounded integrality gap, even when all facilities have the same capacity u and we are allowed to open $(2 - \epsilon)k$ facilities. In the gap instance we have u separate groups, each containing $u + 1$ clients and 2 facilities, and we are allowed to open $k = u + 1$ facilities. The fractional solution can open $1 + 1/u$ facilities in each group and use them to serve the $(1 + 1/u)u = u + 1$ clients in the group. (This can be achieved by, for example, setting $y_i = 1/2 + 1/(2u)$ for each of the two facilities i and $x_{i,j} = 1/2$ every facility i and client j in the group.) In an integral solution, there is a group in which at most one facility is open, even if we are allowed to open $2u - 1 = 2k - 3$ facilities. Some client in this group must be connected to a facility outside the group. Thus, if the groups are far away from each other, the integrality gap is unbounded.

The gap instance suggested the following bad situation. There is an isolate group B of facilities, with some nearby clients. The fractional solution uses y_B open facilities to serve the nearby clients; however serving these clients integrally requires at least $\lceil y_B \rceil$ facilities. So if y_B is a small integer plus a tiny fractional number, then $\lceil y_B \rceil / y_B > 1 + \epsilon$. So, we can not afford to open $\lceil y_B \rceil$ facilities inside B .

This motivates the following idea to handle the isolated group B . Let $\ell_1 = \Theta(1/\epsilon)$ be large enough. If $y_B \geq \ell_1$, then we will have $\lceil y_B \rceil / y_B \leq 1 + \epsilon$. So, the bad case happens only if $y_B \leq \ell_1$. If we know that the optimum solution opens at most ℓ_1 facilities in B , then we can afford to have configuration variables such as $z_{S,i,j}^B$, for every $S \subseteq B, |S| \leq \ell_1, i \in S, j \in C$, to indicate the event that the set of open facilities in B is exactly S , and j is connected to i . With some natural constraints on these variables, we can overcome the gap instance. Since we do not know if the optimum solution opens at most ℓ_1 facilities in B , we shall allow S to take value “ \perp ”, which means that S has size more than ℓ_1 . In this case, we do not care what is the set exactly; knowing that the size is more than ℓ_1 is enough.

Now we formally define the configuration LP. Let us fix a set $B \subseteq F$ of facilities. Let $\mathcal{S} = \{S \subseteq B : |S| \leq \ell_1\}$ and $\tilde{\mathcal{S}} = \mathcal{S} \cup \{\perp\}$, where \perp stands for “any subset of B with size more than ℓ_1 ”; for convenience, we also treat \perp as a set such that $i \in \perp$ holds for every $i \in B$. For $S \in \mathcal{S}$, let z_S^B indicate the event that the set of open facilities in B is exactly S and z_\perp^B indicate the event that the number of open facilities in B is more than ℓ_1 .

For every $S \in \tilde{\mathcal{S}}$ and $i \in S$, $z_{S,i}^B$ indicates the event that $z_S^B = 1$ and i is open. (If $i \in B$ but $i \notin S$, then the event will not happen.) Notice that when $i \in S \neq \perp$, we always have $z_{S,i}^B = z_S^B$; we keep both variables to simplify the description of the configuration LP. For every $S \in \tilde{\mathcal{S}}, i \in S$ and client $j \in C$, $z_{S,i,j}^B$ indicates the event that $z_{S,i}^B = 1$ and j is connected to i . In an integral solution, all the above variables are $\{0, 1\}$ variables. The following constraints are valid. To help understand the constraints, it is good to think of $z_{S,i}^B$ as $z_S^B \cdot y_i$ and $z_{S,i,j}^B$ as $z_S^B \cdot x_{i,j}$.

$$\sum_{S \in \tilde{\mathcal{S}}} z_S^B = 1; \quad (6) \quad z_{S,i}^B = z_S^B, \quad \forall S \in \mathcal{S}, i \in S; \quad (10)$$

$$\sum_{S \in \tilde{\mathcal{S}}: i \in S} z_{S,i}^B = y_i, \quad \forall i \in B; \quad (7) \quad \sum_{i \in S} z_{S,i,j}^B \leq z_S^B, \quad \forall S \in \tilde{\mathcal{S}}, j \in C; \quad (11)$$

$$\sum_{S \in \tilde{\mathcal{S}}: i \in S} z_{S,i,j}^B = x_{i,j}, \quad \forall i \in B, j \in C; \quad (8) \quad \sum_{j \in C} z_{S,i,j}^B \leq u_i z_{S,i}^B, \quad \forall S \in \tilde{\mathcal{S}}, i \in S; \quad (12)$$

$$0 \leq z_{S,i,j}^B \leq z_{S,i}^B \leq z_S^B, \quad \forall S \in \tilde{\mathcal{S}}, i \in S, j \in C; \quad (9) \quad \sum_{i \in B} z_{\perp,i}^B \geq \ell_1 z_\perp^B. \quad (13)$$

Constraint (6) says that $z_S^B = 1$ for exactly one $S \in \tilde{\mathcal{S}}$. Constraint (7) says that if i is open then there is exactly one $S \in \tilde{\mathcal{S}}$ such that $z_{S,i}^B = 1$. Constraint (8) says that if j is connected to i then

there is exactly one $S \in \tilde{\mathcal{S}}$ such that $z_{S,i,j}^B = 1$. Constraint (9) is by the definition of the variables. Constraint (10) holds as we mentioned earlier. Constraint (11) says that if $z_S^B = 1$ then j can be connected to at most 1 facility in S . Constraint (12) is the capacity constraint. Finally, Constraint (13) says that if $z_\perp^B = 1$, then at least ℓ_1 facilities in B are open.

If $z_\perp^B = 0$, the above polytope is integral; this can be seen from the integrality of matching polytopes. This is not true if $z_\perp^B > 0$. The variables $\{z_{\perp,i}^B/z_\perp^B\}_{i \in B}$ and $\{z_{\perp,i,j}^B/z_\perp^B\}_{i \in B, j \in C}$ only define a fractional solution to the basic LP for the instance defined by B and C (not all clients in C need to be connected). This is sufficient as the bad case happens only if a few facilities are open in B .

Our configuration LP is obtained from the basic LP, by adding the z variables and Constraints (6) to (13) for every $B \subseteq F$. Fixing a set $B \subseteq F$ and (x, y) , we can check in time $n^{O(1/\epsilon)}$ if there are z variables satisfying Constraints (6) to (13), since the total number of variables and constraints is $n^{O(\ell_1)=O(1/\epsilon)}$. As there are exponential number of sets B , we do not know how to solve this LP.

Instead, we transform the above configuration LP so that it contains only x and y variables. The new LP will have infinite number of constraints. Fix a subset $B \subseteq F$. Constraints (6) to (13) can be written as $Mz \succeq b + M'x + M''y$, where M, M', M'' are some matrices, x, y, z are the column vectors containing all x, y, z variables respectively, and b is a column vector¹. By the duality of linear programming, the set of constraints is feasible for a fixed (x, y) , iff for every vector g such that $g^T M = 0$, we have $g^T (b + M'x + M''y) \leq 0$. Thus, we can convert Constraints (6) to (13) in the following way: for every g such that $g^T M = 0$, we have the constraint $g^T (b + M'x + M''y) \leq 0$. All these constraints are linear in x and y variables. Given a fixed (x, y) for which the system defined by constraints (6) to (13) is infeasible, we can find a vector g such that $g^T M = 0$ and $g^T (b + M'x + M''y) > 0$.

Thus, our final configuration LP contains only x, y variables, but infinite number of constraints². We can apply the standard trick, which has been used in, e.g. [3] and [22]. Given a fractional solution (x, y) to the basic LP relaxation, our rounding algorithm either constructs an integral solution with the desired properties, or outputs a set $B \subseteq F$ such that Constraints (6) to (13) are infeasible. In the latter case, we can find a violated constraint. This allows us to run the ellipsoid method.

3 An $(O(1), O(1))$ -approximation based on the basic LP relaxation

In this section, we describe an $(O(1), O(1))$ -approximation for CKM based on the basic LP relaxation. This result is not new and our constants are worse than those in [1]. The purpose of section is only to set up a starting line for our $\left(1 + \epsilon, O\left(\frac{1}{\epsilon^2} \log \frac{1}{\epsilon}\right)\right)$ -approximation; most of the components in this section will be used in our new algorithm.

After the set of open facilities is decided, the optimum connection assignment from clients to facilities can be computed by solving a minimum cost b -matching instance. Due to the integrality of the matching polytope, we may allow the connections to be fractional. That is, if there is a good fractional assignment, then there is a good integral assignment. So we can use the following framework. Initially there is one unit of demand at each client $j \in C$. During our algorithm, we move demands fractionally inside $F \cup C$, incurring a cost of $xd(i, j)$ for moving x units of demand from i to j . At the end, all the demands are moved to F . If a facility $i \in F$ has α_i units of demand, we open $\lceil \alpha_i / u_i \rceil$ copies of i to satisfy the demand. Our goal is to bound the total moving cost and the number of open facilities.

A standard approach to facility location problems is to partition the facilities into many clusters.

¹We break an equality into two inequalities.

²Since the polytope defined by x, y, z variables has finite number of facets, so does its projection to x, y coordinates. Thus, only a finite number of constraints matter. However, it is hard to define this finite set; and it is not important.

Each cluster contains a set of nearby facilities and the fractional number of open facilities in each cluster is not too small. Each cluster is associated with a center $v \in C$ which we call *client representatives*.

Focus on a fractional solution (x, y) to the basic LP. Let $d_{av}(j) = \sum_{i \in F} x_{i,j} d(i, j)$ be the connection cost of j , for every client $j \in C$. Then the value of the solution (x, y) is $\text{LP} := \sum_{i \in F, j \in C} x_{i,j} d(i, j) = \sum_{j \in C} d_{av}(j)$. For any set $F' \subseteq F$ of facilities and any set $C' \subseteq C$ of clients, we shall let $x_{F', C'} := \sum_{i \in F', j \in C'} x_{i,j}$; we simply write $x_{i, C'}$ for $x_{\{i\}, C'}$ and $x_{F', j}$ for $x_{F', \{j\}}$. For any $F' \subseteq F$, let $y_{F'} := y(F') := \sum_{i \in F'} y_i$.

We shall use R to denote the set of client representatives. Let $R = \emptyset$ initially. Repeat the following process until C becomes empty. We select the client $v \in C$ with the smallest $d_{av}(v)$ and add it to R . We remove all clients j such that $d(j, v) \leq 4d_{av}(j)$ from C (thus, v itself is removed). We shall use v and its derivatives to index representatives, and j and its derivatives to index general clients.

We partition the set F of locations according to their nearest representatives in R . Let $U_v = \emptyset$ for every $v \in R$ initially. For each location $i \in F$, we add i to U_v for the $v \in R$ that is closest to i . Thus, $\{U_v : v \in R\}$ forms a Voronoi diagram of F with R being the centers. For any subset $V \subseteq R$ of representatives, we use $U_V := U(V) := \bigcup_{v \in V} U_v$ to denote the union of Voronoi regions with centers V .

Claim 3.1. *The following statements hold:*

- (3.1a) *for all $v, v' \in R, v \neq v'$, we have $d(v, v') > 4 \max\{d_{av}(v), d_{av}(v')\}$;*
- (3.1b) *for all $j \in C$, there exists $v \in R$, such that $d_{av}(v) \leq d_{av}(j)$ and $d(v, j) \leq 4d_{av}(j)$;*
- (3.1c) *$y(U_v) \geq 1/2$ for every $v \in R$;*
- (3.1d) *for any $v \in R, i \in U_v$ and $j \in C$, we have $d(i, v) \leq d(i, j) + 4d_{av}(j)$.*

We move all demands to client representatives. First, for each client $j \in C$ and $i \in F$, we move $x_{i,j}$ units of demand from j to i . Obviously, the moving cost is exactly LP . After this step, each $i \in F$ has $\sum_{j \in C} x_{i,j} = x_{i,C} \leq u_i y_i$ units of demand. Second, for each $v \in R$ and $i \in U_v$, we move the $x_{i,C}$ units of demand at i to v . Then, all demands are moved to R and a representative $v \in R$ has $\sum_{i \in U_v} x_{i,C} = x_{U_v, C}$ units of demand. Corollary 3.4 bounds the moving cost for the second step.

Definition 3.2. *Let $D_i := \sum_{j \in C} x_{i,j} d(i, j)$ and $D'_i := \sum_{j \in C} x_{i,j} d_{av}(j)$ for every $i \in F$. Let $D_S := \sum_{i \in S} D_i$ and $D'_S := D'(S) := \sum_{i \in S} D'_i$ for every $S \subseteq F$. Obviously $D_F = D'_F = \text{LP}$.*

Thus, we can think of D_i (similarly, D'_i) as a distribution of the total cost LP to the facilities. We shall use D and D' to charge the cost of integral solutions.

Lemma 3.3. *For every $v \in R$, we have $\sum_{i \in U_v} x_{i,C} d(i, v) \leq D(U_v) + 4D'(U_v)$.*

Proof. By Property (3.1d), we have $d(i, v) \leq d(i, j) + 4d_{av}(j)$ for every $i \in U_v$ and $j \in C$. Thus,

$$\sum_{i \in U_v} x_{i,C} d(i, v) \leq \sum_{i \in U_v, j \in C} x_{i,j} (d(i, j) + 4d_{av}(j)) = \sum_{i \in U_v} (D_i + 4D'_i) = D(U_v) + 4D'(U_v). \quad \square$$

Adding the above inequality for all $v \in R$ gives the following corollary.

Corollary 3.4. $\sum_{v \in R, i \in U_v} x_{i,C} d(i, v) \leq D_F + 4D'_F = 5\text{LP}$.

So, moving demands from F to their respective centers incurs a cost of at most 5LP . Focus on a representative $v \in R$; it has $x_{U_v, C}$ units of demand after the moving. For each $i \in U_v$, we shall move α_i units of demand from v to i . Consider the following LP with variables $\{\alpha_i\}_{i \in U_v}$: minimize $\sum_{i \in U_v} \alpha_i d(i, v)$ subject to $\sum_{i \in U_v} \alpha_i / u_i \leq y(U_v)$, $\sum_{i \in U_v} \alpha_i = x_{U_v, C}$ and $\alpha_i \in [0, u_i]$ for every $i \in U_v$. The value of the LP is at most $\sum_{i \in U_v} x_{i,C} d(i, v)$ which can be achieved by setting $\alpha_i = x_{i,C}$ for every $i \in U_v$. We select a vertex solution $\{\alpha_i^*\}_{i \in U_v}$ for the LP. Then all but at most 2 facilities $i \in U_v$ have $\alpha_i^* \in \{0, u_i\}$. Each $i \in U_v$ gets α_i^* units of demands from v ; the moving cost is $\sum_{i \in U_v} \alpha_i^* d(i, v) \leq \sum_{i \in U_v} x_{i,C} d(i, v) \leq D(U_v) + 4D'(U_v)$ by Lemma 3.3. We open $\sum_{i \in U_v} \lceil \alpha_i^* / u_i \rceil \leq \lfloor \sum_{i \in U_v} \alpha_i^* / u_i + 2 \rfloor \leq \lfloor y(U_v) \rfloor + 2$ facilities

in U_v . The total moving cost over all $v \in R$ in this step is $\sum_{v \in R} (D(U_v) + 4D'(U_v)) = 5\text{LP}$. Since $y(U_v) \geq 1/2$ for every $v \in R$ by Property (3.1c), we have $\frac{\lfloor y(U_v) \rfloor + 2}{y(U_v)} \leq 4$, implying that at most $4k$ facilities are open. This gives a $(4, 11)$ -approximation for CKM.

4 Rounding a fractional solution to the configuration LP

In this section, we show how to round a fractional solution to the configuration LP to obtain our $(1 + \epsilon, O(\frac{1}{\epsilon^2} \log \frac{1}{\epsilon}))$ -approximation. To be more accurate, given a fractional solution (x, y) to the basic LP, the rounding algorithm either succeeds, or finds a set $B \subseteq F$ for which Constraints (6) to (13) is infeasible. $d_{av}(j), \text{LP}, x_{F', C'}, y_{F'} = y(F'), D$ and D' are defined in the same way as in Section 3. We also construct the set R of client representatives and the clustering $\{U_v\}_{v \in R}$ as in Section 3. We let $d(A, B) := \min_{i \in A, j \in B} d(i, j)$ denote the minimum distance between A and B , for any $A, B \subseteq F \cup C$; we simply use $d(i, B)$ for $d(\{i\}, B)$. In this section, $\ell = \Theta(1/\epsilon)$ is a large enough integer, whose value will be given later, $\ell_1 = 2\ell + 2$ and $\ell_2 = \Theta(\ell \log \ell)$ is large enough, given later by Lemma 4.7.

We give an overview of the algorithm. The rounding algorithm in Section 3 opens $\lfloor y(U_v) \rfloor + 2$ facilities inside U_v . If $y(U_v) \geq \Theta(1/\epsilon)$ for every $v \in R$, then the algorithm gives $O(1)$ -approximation with $(1 + \epsilon)k$ open facilities. But we only have $y(U_v) \geq 1/2$. In order to save the number of opening facilities, we follow the framework of [22] that combines the representatives to form bigger groups. We move demands within each group. In each group, we open 2 more facilities than the number given by the fractional solution. To obtain the improved approximation ratio, we use a novel process to partition R into groups, based on coloring the edges of the minimum spanning tree of the metric (R, d) .

If there are no so-called “concentrated sets” in a group, then the moving cost within the group can be charged locally using the D and D' values. For a concentrated set J , we need Constraints (6) to (13) to hold for the set $B = U_J$. We pre-open a set of facilities in U_J and pre-assign a set of clients to these facilities based on the values of the z variables. After the pre-assignment, the moving cost for the remaining demands can be charged locally. We now describe the algorithm in more detail.

Partition R into groups To partition the set R of representatives into groups, we run the classic Kruskal’s algorithm to find the minimum spanning tree MST of the metric (R, d) , and then color the edges in MST using black, grey and white colors. In Kruskal’s algorithm, we maintain the set E_{MST} of edges added to MST so far and a partition \mathcal{J} of R into groups. Initially, we have $E_{\text{MST}} = \emptyset$ and $\mathcal{J} = \{\{v\} : v \in R\}$. The length of an edge $e \in \binom{R}{2}$ is the distance between the two endpoints of e . We sort all edges in $\binom{R}{2}$ according to their lengths, breaking ties arbitrarily. For each pair (u, v) in this order, if u and v are not in the same group in \mathcal{J} , we add the edge (u, v) to E_{MST} and merge the two groups containing u and v respectively.

We now color the edges in E_{MST} . For every $v \in R$, we say the weight of v is $y(U_v)$; so every representative $v \in R$ has weight at least $1/2$ by Property (3.1c). For a subset $J \subseteq R$ of representatives, we say J is large if the weight of J is at least ℓ , i.e., $y(U_J) \geq \ell$; we say J is small otherwise. For any edge $e = (u, v) \in E_{\text{MST}}$, we consider the iteration in Kruskal’s algorithm in which the edge e is added to MST . After the iteration we merged the group J_u containing u and the group J_v containing v into a new group $J_u \cup J_v$. If both J_u and J_v are small, then we call e a black edge. If J_u is small and J_v is big, we call e a grey edge, directed from u to v ; similarly, if J_v is small and J_u is big, e is a grey edge directed from v to u . If both J_u and J_v are big, we say e is a white edge. So, we treat black and white edges as undirected edges and grey edges as directed edges.

We define a black component of MST to be a maximal set of vertices connected by black edges. The following claim is straightforward. It follows from the fact that a black component $J \subseteq R$ appeared as

a group at some iteration of Kruskal's algorithm for computing MST.

Claim 4.1. *Let J be a black component of MST. Then for every black edge (u, v) in $\binom{J}{2}$, we have $d(u, v) \leq d(J, R \setminus J)$.*

We contract all the black edges in MST and remove all the white edges. The resulting graph is a forest Υ of trees. Each node (we use the word “nodes” for vertices in the contracted graph) in Υ is correspondent to a black component, and each edge is a directed grey edge. For every node p in Υ , we use $J_p \subseteq R$ to denote the black component correspondent to the node p . Abusing notations slightly, we define $U_p := U(J_p) = \bigcup_{v \in J_p} U_v$. The weight of the node p is the total weight of representatives in J_p , i.e., $y(U_p)$.

Lemma 4.2. *For any tree $\tau \in \Upsilon$, the following statements are true:*

- (4.2a) τ has a root node r_τ such that all grey edges in τ are directed towards r_τ ;
- (4.2b) J_{r_τ} is big and J_p is small for all other nodes p in τ ;
- (4.2c) in any leaf-to-root path of τ , the lengths of grey edges form a non-increasing sequence;
- (4.2d) for any non-root node p of τ , the length of the grey edge in τ connecting p to its parent is exactly $d(J_p, R \setminus J_p)$;
- (4.2e) for any non-root node p of τ , the length of any black edge in $\binom{J_p}{2}$ is at most $d(J_p, R \setminus J_p)$.

We now break τ into a set \mathcal{T}_τ of edge-disjoint sub-trees using the following greedy algorithm. Consider the deepest node p in τ such that the total weight of all descendant nodes of p (we do not count the weight of p) is at least ℓ . First assume that the node p exists.

Notice that the weight of each node other than the root node r_τ is less than ℓ by Property (4.2b). By our choice of p , the weight of any sub-tree rooted at a child of p is at most $\ell + \ell = 2\ell$. Thus, a simple greedy algorithm can give us a collection of sub-trees rooted at some children of p , with total weight between ℓ and 2ℓ . We build a tree T as follows: take the collection of sub-trees, the node p , as well as the edges connecting the roots of the sub-trees to p . We add the tree T into \mathcal{T}_τ . Let $r_T = p$ be the root node of T .

We remove the collection of sub-trees from τ and repeat the above process to find another tree T . We terminate the process when the node p does not exist. In this case, we add the remaining tree to \mathcal{T}_τ . It is easy to see the following statements about \mathcal{T}_τ .

Lemma 4.3. (4.3a) *Every non-root node p of τ appears in exactly one tree in \mathcal{T}_τ as a non-root.*

(4.3b) *The number of trees in \mathcal{T}_τ is at most $1/\ell$ times the total weight of nodes in τ .*

(4.3c) *Let $T \in \mathcal{T}_\tau$ and \tilde{T} be the tree obtained from T by un-contracting the nodes of T to their correspondent black components. Then \tilde{T} contains at most 8ℓ vertices.*

Proof. Property (4.3a) simply follows from the construction of the \mathcal{T}_τ . Focus on a tree $T = (P, E_T) \in \mathcal{T}_\tau$ rooted at r that is not the last tree added to \mathcal{T}_τ . The total weight of all nodes in $P \setminus r$ is $y(U(P \setminus r)) \in [\ell, 2\ell]$. Let the “contributing set” of the tree T be $P \setminus r$. The total weight all nodes in the last tree added to \mathcal{T}_τ is at least ℓ , since the root node r_τ has weight at least ℓ . Let the “contributing set” of the last tree T be P . Since the contributing sets are disjoint and each contributing set has weight at least ℓ , the number of trees in \mathcal{T}_τ is at most $1/\ell$ times the total weight of τ . We proved Property (4.3b).

Focus on a tree $T = (P, E_T) \in \mathcal{T}_\tau$ such that the root of T is not r_τ . If we uncontract all nodes in P to their respective black components, then the number of vertices in the resulting tree is at most 6ℓ . This holds since the total weight of all nodes in P is at most 3ℓ and each vertex v has weight $y(U_v) \geq 1/2$. Now suppose the root of T is r_τ . The root node r_τ either has weight at most 2ℓ , or J_{r_τ} is a singleton.

In either case, we can bound the number of vertices in \tilde{T} by $(2\ell + 2\ell)/(1/2) = 8\ell$. So Property (4.3c) holds. \square

The sub-trees in \mathcal{T}_τ for all trees $\tau \in v$ gave us a partition of R . For every tree $\tau \in \Upsilon$ and every $T = (P, E_T) \in \mathcal{T}_\tau$ with root r , there is a group $J_{P \setminus r}$ in the partition. Also, for any root r_τ of $\tau \in \Upsilon$, there is a group J_{r_τ} in the partition. Our algorithm handles each group separately.

Pre-opening facilities and pre-assigning clients In order to bound the moving cost within each group, we need to handle the so-called “concentrated sets”. Before defining concentrated sets, we need to define an important function of sets of representatives:

Definition 4.4. Let $\pi(J) = \sum_{j \in C} x_{U(J),j}(1 - x_{U(J),j})$, for every $J \subseteq R$.

The next lemma shows the importance of the function $\pi(J)$.

Lemma 4.5. Given any non-trivial subset $J \subseteq R$ of representatives, we have

$$d(J, R \setminus J)\pi(J) \leq 4D(U_J) + 10D'(U_J).$$

Notice that in an isolated group $B = U_J$ in the gap instance, each client is either completely served by B , or completely served by $F \setminus B$. Thus $\pi(J)$ is 0. So the inequality holds no matter how big $d(J, R \setminus J)$ is. In some sense, $\pi(J)$ measures how many clients are both served by U_J and $F \setminus U_J$. The bigger $\pi(J)$, the smaller $d(J, R \setminus J)$ is.

Definition 4.6. A set $J \subseteq R$ of representatives is said to be concentrated if $\pi(J) \leq x_{U(J),C}/\ell_2$.

Recall that $x_{U(J),C}$ is the total demand in U_J after all demands are moved to the representatives using the algorithm in Section 3 and $\ell_2 = \Theta(\ell \log \ell)$ is a large enough number. Thus, according to Lemma 4.5, if J is not concentrated, we can use $D(U_J) + D'(U_J)$ to charge the cost for moving all the $x_{U(J),C}$ units of demand out of J , provided that the moving distance is not too big compared to $d(J, R \setminus J)$. If J is concentrated, the amount of demand that is moved out of J must be comparable to $\pi(J)$. To achieve this goal, we will use the following lemma to pre-assign some clients C' so that the remaining demands $x_{U_J, C \setminus C'}$ inside U_J is comparable to $\pi(J)$.

Lemma 4.7. If $\ell_2 = O(\ell \log \ell)$ is large enough then the following is true. Let $J \subseteq R$ be a concentrated set and $B = U_J$ satisfies $y_B \leq 2\ell$. Moreover, Constraints (6) to (13) are satisfied for B . Then, we can pre-open a set $S \subseteq B$ of facilities and pre-assign a set $C' \subseteq C$ of clients to S such that

(4.7a) each facility $i \in S$ is pre-assigned at most u_i clients;

(4.7b) $x_{B, C \setminus C'} \leq \ell_2 \pi(J)$;

(4.7c) $\frac{x_{B, C \setminus C'}}{x_{B, C}} y_B + |S| \leq \left(1 + \frac{1}{\ell}\right) y_B$;

(4.7d) the cost for the pre-assignment is at most $\ell_2 D_B$.

Property (4.7b) says that the remaining demand in $B = U_J$ is small. Property (4.7d) says that the cost for the pre-assignment can be charged locally using D_B . Property (4.7c) says that even after we pre-opened the facilities in S , we can still afford to open $\frac{x_{B, C \setminus C'}}{x_{B, C}} y_B$ facilities. Notice that if J is not concentrated, we can simply take $S = \emptyset$ and $C = \emptyset$ to satisfy all the properties.

We defer the formal proof of Lemma 4.7 to Appendix A and give the key ideas here, assuming $\pi(J) = 0$ and $z_\perp^B = 0$. The fractional vector on variables $\{x_{i,j}\}_{i \in B, j \in C}$ and $\{y_i\}_{i \in B}$ can be expressed as a convex combination of valid integral vectors. That is, we can randomly open a set $S \subseteq B$ of facilities and connect some facilities C' to facilities in S such that: (i) each facility i is open with probability y_i ; (ii) each client j is connected with probability $x_{B,j}$ and (iii) each facility i is connected by at most u_i clients. $\pi(J) = 0$ implies that every client j has either $x_{B,j} = 0$ or $x_{B,j} = 1$. If $x_{B,j} = 1$, $j \in C'$

always holds. Thus, we always have $x_{B,C \setminus C'} = 0$, which implies Properties (4.7b) and Property (4.7c). The expected assignment cost will be at most D_B . We condition on the event that $|S| \leq (1 + 1/\ell)y_B$, which happens with probability at least $\Omega(1/\ell)$. So, under this condition, the expected connection cost will be $O(\ell)D_B$, implying Property (4.7d). When $\pi(J) > 0$, we use a smooth version of this proof. As we only conditioned on the event that $|S| \leq (1 + 1/\ell)y_B \leq \ell_1$, $z_\perp^B > 0$ is not an issue. In the proof, we group the sets in \mathcal{S} into $O(\log \ell)$ groups; this is the reason that we require $\ell_2 = \Theta(\ell \log \ell)$.

With Lemma 4.7, we can now pre-open some facilities, and pre-assign some clients to these pre-opened facilities. We handle each tree $\tau \in \Upsilon$ separately. For every node p in τ other than the root r_τ , the weight of p is at most ℓ . For the root r_τ , either the weight of r_τ is at most 2ℓ , or J_{r_τ} contains only one vertex. So, we apply the following procedure to p if p is not the root, or if p is the root and the weight of p is at most 2ℓ . If J_p is concentrated, we check if Constraints (6) to (13) are satisfied for $B = U_p$. If not, we return the set $B = U_p$. Otherwise, we apply Lemma 4.7 to the set J_p to pre-open some facilities and pre-assign some clients. Notice that the total cost for pre-assignment is at most $\ell_2 D_F = \ell_2 \text{LP}$, as the sets J for which we apply Lemma 4.7 are disjoint.

As the pre-assignment processes for all nodes p are done independently, it is possible that a client is pre-assigned more than once. This is not an issue as we only over-estimated the cost for the pre-assignment. Let \tilde{C} be the set of clients that are never pre-assigned.

Moving Demands We only need to focus on clients in \tilde{C} now. Each client $j \in \tilde{C}$ initially has one unit of demand. We need to move all demands to facilities. First, for every $j \in \tilde{C}$ and $i \in F$, we move $x_{i,j}$ units of demand from j to i . The moving cost is $\sum_{j \in \tilde{C}} d_{\text{av}}(j) \leq \text{LP}$. After this step, each facility $i \in F$ has $x_{i,\tilde{C}}$ units of demand.

Let us focus on a tree τ in the forest Υ and a tree $T = (P, E) \in \mathcal{T}_\tau$ with root node r . P is the set of nodes in T and E is the set of grey edges. Recall that $U_{P \setminus r} = \bigcup_{p \in P \setminus r} U_p$. We move the demands within $U_{P \setminus r}$. The moving process is simple. Let $v^* \in J_r$ be an arbitrary representative in J_r . We first move all demands in $U_{P \setminus r}$ to v^* , then we move the demand in v^* to $U_{P \setminus r}$ according to some distribution.

In order to find the distribution, we shall use α_i to denote the amount of demand we shall move to facility i , for any $i \in U_{P \setminus r}$. Let $\alpha(S) := \sum_{i \in S} \alpha_i$ for every $S \subseteq U_{P \setminus r}$. Let t be the number of pre-opened facilities in $U_{P \setminus r}$. As we are considering the soft-capacitated case, we can open the facility i even if i is pre-opened. We solve the following LP to obtain $\{\alpha_i\}_{i \in U_{P \setminus r}}$:

$$\begin{aligned} \min \quad & \sum_{i \in U_{P \setminus r}} \alpha_i d(i, v^*) \quad \text{s.t.} \\ & \alpha_i \in [0, u_i], \forall i \in U_{P \setminus r}; \quad \alpha(U_{P \setminus r}) = x_{U_{P \setminus r}, \tilde{C}}; \quad \sum_{i \in U_{P \setminus r}} \frac{\alpha_i}{u_i} + t \leq (1 + \frac{1}{\ell})y(U_{P \setminus r}). \end{aligned} \quad (14)$$

The objective function of LP(14) is the moving cost. The first constraint requires that we move at most u_i units of demand to i , the second constraint says the total demand is $x_{U_{P \setminus r}, \tilde{C}}$, and the last constraint bounds the total number of open facilities, including the t pre-opened facilities. We give a valid solution $\{\tilde{\alpha}_i\}_{i \in U_{P \setminus r}}$ for the LP: for every $p \in P \setminus r$ and $i \in U_p$, we let $\tilde{\alpha}_i = \frac{x_{U_p, \tilde{C}}}{x_{U_p, C}} x_{i,C}$.

Claim 4.8. $\{\tilde{\alpha}_i\}_{i \in U_{P \setminus r}}$ satisfies the constraints of LP(14).

We find a vertex solution $(\alpha_i^*)_{i \in U_{P \setminus r}}$ of LP(14). So, for all but at most 2 facilities $i \in U_{P \setminus r}$, we have $\alpha_i^* \in \{0, u_i\}$. $\{\alpha_i^*\}_{i \in U_{P \setminus r}}$ satisfies the constraints of LP(14), and $\sum_{i \in U_{P \setminus r}} \alpha_i^* d(i, v^*) \leq \sum_{i \in U_{P \setminus r}} \tilde{\alpha}_i d(i, v^*)$.

We open a facility at i if $\alpha_i^* > 0$. The total number of open facilities, including the pre-opened facilities, is at most $\sum_{i \in U_{P \setminus r}} \lceil \alpha_i^* / u_i \rceil + t \leq \sum_{i \in U_{P \setminus r}} \alpha_i^* / u_i + 2 + t \leq (1 + 1/\ell)y(U_{P \setminus r}) + 2$.

The total moving cost is at most $\sum_{i \in U_{P \setminus r}} (x_{i,\tilde{C}} + \alpha_i^*) d(i, v^*) \leq \sum_{i \in U_{P \setminus r}} (x_{i,\tilde{C}} + \tilde{\alpha}_i) d(i, v^*)$. Focus on

a node $p \in P \setminus r, v \in J_p$ and $i \in U_v$. We have

$$\begin{aligned} (x_{i,\tilde{C}} + \tilde{\alpha}_i)d(i, v^*) &\leq (x_{i,\tilde{C}} + \tilde{\alpha}_i)(d(i, v) + d(v, v^*)) \leq 2x_{i,C}d(i, v) + \left(x_{i,\tilde{C}} + \frac{x_{U_p,\tilde{C}}}{x_{U_p,C}}x_{i,C}\right)d(v, v^*) \\ &\leq 2x_{i,C}d(i, v) + 8\ell\left(x_{i,\tilde{C}} + \frac{x_{U_p,\tilde{C}}}{x_{U_p,C}}x_{i,C}\right)d(J_p, R \setminus J_p). \end{aligned}$$

The first inequality is by triangle inequality. The second inequality used $x_{i,\tilde{C}} \leq x_{i,C}$ and $\tilde{\alpha}_i \leq x_{i,C}$. By Properties (4.2c), (4.2d) and (4.2e), all edges in the path from v to v^* in MST have length at most $d(J_p, R \setminus J_p)$. By Property (4.3c), we have $d(v, v^*) \leq 8\ell d(J_p, R \setminus J_p)$, implying the third inequality.

Fix $p \in P \setminus r$. We sum up the above inequality over all $v \in J_p$ and $i \in U_v$. The first term becomes

$$2 \sum_{v \in J_p, i \in U_v} x_{i,C}d(i, v) \leq 2 \sum_{v \in J_p} (D(U_v) + 4D'(U_v)) = O(1)(D(U_p) + D'(U_p)),$$

by Lemma 3.3. The second term becomes

$$8\ell \sum_{i \in U_p} \left(x_{i,\tilde{C}} + \frac{x_{U_p,\tilde{C}}}{x_{U_p,C}}x_{i,C}\right)d(J_p, R \setminus J_p) = O(\ell)x_{U_p,\tilde{C}}d(J_p, R \setminus J_p).$$

If J_p is not concentrated, then $x_{U_p,\tilde{C}} \leq x_{U_p,C} \leq \ell_2\pi(J_p)$, by the definition of concentrated sets. If J_p is concentrated, then $x_{U_p,\tilde{C}} \leq \ell_2\pi(J_p)$ by Property (4.7b). So, $x_{U_p,\tilde{C}} \leq \ell_2\pi(J_p)$ always holds. The above quantity is at most $O(\ell\ell_2)\pi(J_p)d(J_p, R \setminus J_p) \leq O(\ell\ell_2)(D(U_p) + D'(U_p))$ by Lemma 4.5.

So, we have $\sum_{i \in U_p} (x_{i,\tilde{C}} + \tilde{\alpha}_i)d(i, v^*) \leq O(\ell\ell_2)(D(U_p) + D'(U_p))$. Summing up over all $p \in P \setminus r$, we obtain $\sum_{i \in U_{P \setminus r}} (x_{i,\tilde{C}} + \tilde{\alpha}_i)d(i, v^*) \leq O(\ell\ell_2)(D(U_{P \setminus r}) + D'(U_{P \setminus r}))$.

For a tree $\tau \in \Upsilon$ with root $r = r_\tau$, we also need to move demands within the facilities in U_r . Observe that all black edges in $\binom{J_r}{2}$ have length at most $d(J_r, R \setminus J_r)$ and J_r has at most 4ℓ vertices. Use exactly the same argument as above we can bound the number of open facilities (including the possible pre-opened facilities) in U_r by $(1 + \frac{1}{\ell})y(U_r) + 2$ and the moving cost by $O(\ell\ell_2)(D(U_r) + D'(U_r))$. That is, we let v^* be an arbitrary vertex in J_r . We move all demands in U_r first to v^* ; then we move the demand in v^* to facilities U_r , according to the distribution $\{\alpha_i^*\}_{i \in U_r}$ obtained by solving LP(14), with $U_{P \setminus r}$ replaced by U_r . All the above inequalities hold with $P \setminus r$ replaced with r .

Taking all trees $T \in \mathcal{T}_\tau$ into consideration, the total number of open facilities in U_{P_τ} is at most $(1 + 1/\ell)y(U_{P_\tau}) + 2(|\mathcal{T}_\tau| + 1)$, where P_τ is the set of nodes in τ . By Property (4.3b), this is at most $(1 + 1/\ell)y(U_{P_\tau}) + 2(y(U_{P_\tau})/\ell + 1) \leq (1 + 5/\ell)y(U_{P_\tau})$, since $y(U_{P_\tau}) \geq \ell$. The total moving cost is at most $O(\ell\ell_2)(D(U_{P_\tau}) + D'(U_{P_\tau}))$. Taking all trees $\tau \in \Upsilon$ into consideration, the total number of open facilities is at most $(1 + 5/\ell)y_F \leq (1 + 5/\ell)k$. The total moving cost, is $O(\ell\ell_2)(D(F) + D'(F)) = O(\ell\ell_2)\text{LP}$. Setting $\ell = \lceil 5/\epsilon \rceil$, we obtain our $\left(1 + \epsilon, O\left(\frac{1}{\epsilon^2} \log \frac{1}{\epsilon}\right)\right)$ -approximation for CKM. Due to the pre-opening, a facility may be opened twice in our solution.

5 Discussion

In this paper, we proposed a $\left(1 + \epsilon, O\left(\frac{1}{\epsilon^2} \log \frac{1}{\epsilon}\right)\right)$ -approximation for CKM. We introduced a novel configuration LP for the problem which has small integrality gap with $(1 + \epsilon)k$ open facilities.

There are some open problems related to our result. Our algorithm opens at most 2 copies of each facility. Can we reduce the number of copies for each facility to 1 so that we can extend the result to hard CKM? Can we get constant approximation for CKM with $(1 + \epsilon)$ -violation on the capacity constraints? Finally, can we get a true constant approximation for CKM? The problem is open even for a very special case: all facilities have the same capacity u , the number of clients is exactly $n = ku$, and $F = C$ (which can be assumed w.l.o.g by [22]).

References

- [1] Karen Aardal, Pieter van den Berg, Dion Gijswijt, and Shanfei Li. Approximation algorithms for hard capacitated k -facility location problems. *CoRR*, abs/1311.4759v4, 2014.
- [2] Ankit Aggarwal, L. Anand, Manisha Bansal, Naveen Garg, Neelima Gupta, Shubham Gupta, and Surabhi Jain. A 3-approximation for facility location with uniform capacities. In *Proceedings of the 14th International Conference on Integer Programming and Combinatorial Optimization, IPCO'10*, pages 149–162, Berlin, Heidelberg, 2010. Springer-Verlag.
- [3] Hyung-Chan An, Mohit Singh, and Ola Svensson. LP-based algorithms for capacitated facility location. In *Proceedings of the 55th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2014*.
- [4] V. Arya, N. Garg, R. Khandekar, A. Meyerson, K. Munagala, and V. Pandit. Local search heuristic for k -median and facility location problems. In *Proceedings of the thirty-third annual ACM symposium on Theory of computing, STOC '01*, pages 21–29, New York, NY, USA, 2001. ACM.
- [5] Manisha Bansal, Naveen Garg, and Neelima Gupta. A 5-approximation for capacitated facility location. In *Proceedings of the 20th Annual European Conference on Algorithms, ESA'12*, pages 133–144, Berlin, Heidelberg, 2012. Springer-Verlag.
- [6] J. Byrka. An optimal bifactor approximation algorithm for the metric uncapacitated facility location problem. In *APPROX '07/RANDOM '07: Proceedings of the 10th International Workshop on Approximation and the 11th International Workshop on Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 29–43, Berlin, Heidelberg, 2007. Springer-Verlag.
- [7] Jaroslaw Byrka, Krzysztof Fleszar, Bartosz Rybicki, and Joachim Spoerhase. Bi-factor approximation algorithms for hard capacitated k -median problems. In *Proceedings of the 26th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2015)*.
- [8] Jaroslaw Byrka, Thomas Pensyl, Bartosz Rybicki, Aravind Srinivasan, and Khoa Trinh. An improved approximation for k -median, and positive correlation in budgeted optimization. In *Proceedings of the 26th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2015)*.
- [9] Robert D. Carr, Lisa K. Fleischer, Vitus J. Leung, and Cynthia A. Phillips. Strengthening integrality gaps for capacitated network design and covering problems. In *Proceedings of the Eleventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '00*, pages 106–115, Philadelphia, PA, USA, 2000. Society for Industrial and Applied Mathematics.
- [10] M. Charikar and S. Guha. Improved combinatorial algorithms for the facility location and k -median problems. In *In Proceedings of the 40th Annual IEEE Symposium on Foundations of Computer Science*, pages 378–388, 1999.
- [11] M. Charikar, S. Guha, E. Tardos, and D. B. Shmoys. A constant-factor approximation algorithm for the k -median problem (extended abstract). In *Proceedings of the thirty-first annual ACM symposium on Theory of computing, STOC '99*, pages 1–10, New York, NY, USA, 1999. ACM.
- [12] F. A. Chudak and D. B. Shmoys. Improved approximation algorithms for the uncapacitated facility location problem. *SIAM J. Comput.*, 33(1):1–25, 2004.

- [13] Fabian A. Chudak and David P. Williamson. Improved approximation algorithms for capacitated facility location problems. *Math. Program.*, 102(2):207–222, March 2005.
- [14] Julia Chuzhoy and Yuval Rabani. Approximating k-median with non-uniform capacities. In *SODA '05*, pages 952–958, 2005.
- [15] Samuel Fiorini, Serge Massar, Sebastian Pokutta, Hans Raj Tiwary, and Ronald de Wolf. Linear vs. semidefinite extended formulations: Exponential separation and strong lower bounds. In *Proceedings of the Forty-fourth Annual ACM Symposium on Theory of Computing*, STOC '12, pages 95–106, New York, NY, USA, 2012. ACM.
- [16] S Guha and S Khuller. Greedy strikes back: Improved facility location algorithms. In *Journal of Algorithms*, pages 649–657, 1998.
- [17] K. Jain, M. Mahdian, E. Markakis, A. Saberi, and V. V. Vazirani. Greedy facility location algorithms analyzed using dual fitting with factor-revealing LP. *J. ACM*, 50:795–824, November 2003.
- [18] K. Jain, M. Mahdian, and A. Saberi. A new greedy approach for facility location problems. In *Proceedings of the thirty-fourth annual ACM symposium on Theory of computing*, STOC '02, pages 731–740, New York, NY, USA, 2002. ACM.
- [19] K Jain and V. V. Vazirani. Approximation algorithms for metric facility location and k-median problems using the primal-dual schema and Lagrangian relaxation. *J. ACM*, 48(2):274–296, 2001.
- [20] M. R. Korupolu, C. G. Plaxton, and R. Rajaraman. Analysis of a local search heuristic for facility location problems. In *Proceedings of the ninth annual ACM-SIAM symposium on Discrete algorithms*, SODA '98, pages 1–10, Philadelphia, PA, USA, 1998. Society for Industrial and Applied Mathematics.
- [21] Shanfei Li. An improved approximation algorithm for the hard uniform capacitated k-median problem. In *APPROX '14/RANDOM '14: Proceedings of the 17th International Workshop on Combinatorial Optimization Problems and the 18th International Workshop on Randomization and Computation*, APPROX '14/RANDOM '14, 2014.
- [22] Shi Li. On uniform capacitated k-median beyond the natural LP relaxation. In *Proceedings of the 26th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2015)*.
- [23] Shi Li. A 1.488 approximation algorithm for the uncapacitated facility location problem. In *Automata, Languages and Programming - 38th International Colloquium (ICALP)*, pages 77–88, 2011.
- [24] Shi Li and Ola Svensson. Approximating k-median via pseudo-approximation. In *Proceedings of the Forty-fifth Annual ACM Symposium on Theory of Computing*, STOC '13, pages 901–910, New York, NY, USA, 2013. ACM.
- [25] J. Lin and J. S. Vitter. Approximation algorithms for geometric median problems. *Inf. Process. Lett.*, 44:245–249, December 1992.
- [26] M. Mahdian, Y. Ye, and J. Zhang. Approximation algorithms for metric facility location problems. *SIAM J. Comput.*, 36(2):411–432, 2006.

- [27] D. B. Shmoys, E. Tardos, and K. Aardal. Approximation algorithms for facility location problems (extended abstract). In *STOC '97: Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*, pages 265–274, New York, NY, USA, 1997. ACM.
- [28] Mihalis Yannakakis. Expressing combinatorial optimization problems by linear programs. *Journal of Computer and System Sciences*, 43(3):441 – 466, 1991.
- [29] Jiawei Zhang, Bo Chen, and Yinyu Ye. A multiexchange local search algorithm for the capacitated facility location problem. *Math. Oper. Res.*, 30(2):389–403, May 2005.

A Omitted proofs

A.1 Proof of Claim 3.1

Proof. First consider Property (3.1a). Assume $d_{av}(v) \leq d_{av}(v')$. When we add v to R , we remove all clients j satisfying $d(v, j) \leq 4d_{av}(j)$ from C . Thus, v' can not be added to R later. For Property (3.1b), just consider the iteration in which j is removed from C . The representative v added to R in the iteration satisfy the property. Then consider Property (3.1c). By Property (3.1a), we have $B := \{i \in F : d(i, v) \leq 2d_{av}(v)\} \subseteq U_v$. Since $d_{av}(v) = \sum_{i \in F} x_{i,v} d(i, v)$ and $\sum_{i \in F} x_{i,v} = 1$, we have $d_{av}(v) \geq (1 - x_{B,v})2d_{av}(v)$, implying $y(U_v) \geq y_B \geq x_{B,v} \geq 1 - \frac{1}{2}$, due to Constraint (3).

Finally, consider Property (3.1d). By Property (3.1b), there is a client $v' \in R$ such that $d_{av}(v') \leq d_{av}(j)$ and $d(v', j) \leq 4d_{av}(j)$. Notice that $d(i, v) \leq d(i, v')$ since $v' \in R$ and i was added to U_v . Thus, $d(i, v) \leq d(i, v') \leq d(i, j) + d(j, v') \leq d(i, j) + 4d_{av}(j)$. \square

A.2 Proof of Claim 4.1

Proof. We only need to prove that all the black edges in $\binom{J}{2}$ are considered before all the edges in $J \times (R \setminus J)$ in the Kruskal's algorithm. Assume otherwise. Consider the first edge e in $J \times (R \setminus J)$ we considered. Before this iteration, J is not connected yet. Then we add e to the minimum spanning tree; since J is a black component, e is gray or white. In either case, the new group J' formed by adding e will have weight more than ℓ . This implies all edges in $J' \times (R \setminus J')$ added later to the MST are not black. Moreover, $J \setminus J'$, $J' \setminus J$ and $J \cap J'$ are all non-empty. This contradicts the fact that J is a black component. \square

A.3 Proof of Lemma 4.2

Proof. Focus on any small black component J in MST. By Claim 4.1, it is a group at some iteration of the Kruskal's algorithm. Consider the first iteration that we add an edge in $J \times (R \setminus J)$ to the MST. This edge can not be white because J is small; the edge can not be black since J is a (maximum) black component. Thus, the edge must be a gray edge in MST, directed from J to some other black component.

If for a group $J' \in \mathcal{J}$ at some iteration of Kruskal's algorithm, $\binom{J'}{2}$ contains a gray or white edge, then J' is big. We can only add white edges between two big groups. Let $\tilde{\tau}$ be the tree τ obtained by un-contract all the nodes back to the original black component. The growth of the tree $\tilde{\tau}$ in the Kruskal's algorithm must be as follows. First, a grey edge is added between two black components, one of them is big and the other is small. We define the root node r_τ of τ to be the node correspondent to the big component. At each time, we add a new small black component J to the existing tree via a grey edge with head in J . (During this process, white edges incident to the existing tree $\tilde{\tau}$ may be

added.) So, the tree τ is a rooted tree with grey edges, where all edges are directed towards the root. This proves Property (4.2a) and (4.2b). By the order we add the grey edges, we have Property (4.2c). For each small black component J , the first grey edge in $J \times (R \setminus J)$ is the grey edge between J is its parent component. Thus, we have Property (4.2d). This edge is added after the black component J is formed; thus we have Property (4.2e). \square

A.4 Proof of Lemma 4.5

Proof. Let $B = U_J$. For every $i \in B, j \in C$, we have $d(i, J) \leq d(i, j) + 4d_{av}(j)$ by Property (3.1d) in Claim 3.1 and the fact that $i \in U_v$ for some $v \in J$. Thus,

$$\begin{aligned} d(J, R \setminus J)\pi(J) &= d(J, R \setminus J) \sum_{j \in C} x_{B,j}(1 - x_{B,j}) = d(J, R \setminus J) \sum_{j \in C, i \in B, i' \in F \setminus B} x_{i,j}x_{i',j} \\ &\leq \sum_{i \in B, j \in C, i' \in F \setminus B} x_{i,j}x_{i',j} \cdot 2d(i', J) \leq 2 \sum_{i \in B, j \in C} x_{i,j} \sum_{i' \in F} x_{i',j} (d(i', j) + d(j, i) + d(i, J)) \\ &= 2 \sum_{i \in B, j \in C} x_{i,j} (d_{av}(j) + d(j, i) + d(i, J)) \leq 2 \sum_{i \in B, j \in C} x_{i,j} (2d(i, j) + 5d_{av}(j)) \\ &= 2 \sum_{i \in B} (2D_i + 5D'_i) = 4D(U_J) + 10D'(U_J). \end{aligned}$$

In the above sequence, the first inequality is by $d(J, R \setminus J) \leq 2d(i', J)$ for any $i' \in F \setminus B = U_{R \setminus J}$: $d(i', R \setminus J) \leq d(i', J)$ implies $d(R \setminus J, J) \leq d(R \setminus J, i') + d(i', J) \leq 2d(i', J)$. The second inequality is by triangle inequality and the third one is by $d(i, J) \leq d(i, j) + 4d_{av}(j)$. All the equalities are by simple manipulations of notations. \square

A.5 Pre-assignment of clients: proof of Lemma 4.7

Proof. Let $Y = (1 + 1/\ell)y_B$. For any $S \in \mathcal{S}$ such that $|S| \leq Y$, we give S a rank. If $Y - |S| < 1$, then let $\text{rank}(S) = 0$. Otherwise, if $Y - |S| \in [2^{t-1}, 2^t]$ for some integer $t \geq 1$, then we let $\text{rank}(S) = t$. So, the rank of S is an integer between 0 and $\delta - 1 := \lfloor \log Y \rfloor + 1 = O(\log \ell)$.

We take the assignment of z variables that satisfies Constraints (6) to (13) for $B = U_J$. We have $\sum_{S \in \mathcal{S}} z_S^B |S| + z_\perp^B \ell_1 \leq y_B$ and $\sum_{S \in \tilde{\mathcal{S}}} z_S^B = 1$. So, $\sum_{S \in \mathcal{S}} z_S^B (Y - |S|) + z_\perp^B (Y - \ell_1) \geq Y - y_B = y_B/\ell$. This implies $\sum_{S \in \mathcal{S}: |S| < Y} z_S^B (Y - |S|) \geq y_B/\ell$ as $Y \leq 2\ell(1 + 1/\ell) = \ell_1$. Define a function $f : \{0, 1, 2, \dots, \delta\} \rightarrow \mathbb{R}$ as $f(0) = Y - \lfloor Y \rfloor$ and $f(t) = 2^t$ if $t \geq 1$. If a set S has rank $t \geq 1$, then $Y - |S| \in [f(t)/2, f(t)]$. If a set S has rank 0, then $|S| = \lfloor Y \rfloor$ and $Y - |S| = Y - \lfloor Y \rfloor = f(0)$. Thus, we always have $Y - |S| \in [f(\text{rank}(S))/2, f(\text{rank}(S))]$.

$$\sum_{S \in \mathcal{S}: |S| < Y} z_S^B f(\text{rank}(S)) \geq \sum_{S \in \mathcal{S}: |S| < Y} z_S^B (Y - |S|) \geq \frac{y_B}{\ell}.$$

So, there is an integer $t \in [0, \delta - 1]$ such that

$$f(t) \sum_{S \in \mathcal{S}: \text{rank}(S) = t} z_S^B \geq \frac{y_B}{\delta \ell}.$$

Since $\sum_{S \in \tilde{\mathcal{S}}} z_S^B = 1$, the variables $\{z_S^B\}_{S \in \tilde{\mathcal{S}}}$ defines a distribution over $\tilde{\mathcal{S}}$. We randomly select an element $S \in \tilde{\mathcal{S}}$: S is selected with probability z_S^B . Let $* = \{S \in \mathcal{S} : \text{rank}(S) = t\}$; we are interested in

the event that $S \in *$. Let $q = \Pr[S \in *]$. Then

$$q = \sum_{S \in *} z_S^B \geq \frac{y_B}{\delta \ell f(t)} \geq \Omega\left(\frac{1}{\ell \log \ell}\right), \quad \text{as } f(t) \leq 2Y \leq 3y_B.$$

Now assume we have selected a set $S \in *$. For every facility $i \in S$ and every client $j \in C$, let $w_{i,j} := z_{S,i,j}^B / z_{S,i}^B = z_{S,i,j}^B / z_S^B$, by Constraint (10). Notice that $w_{i,j} \in [0, 1]$ by Constraint (9), $\sum_{i \in S} w_{i,j} \leq 1$ for every $j \in C$ by Constraint (11), and $\sum_{j \in C} w_{i,j} \leq u_i$ for every $i \in S$ by Constraint (12). Thus, w is a fractional matching between S and C such that every facility in $i \in S$ is matched at most u_i times and every client in C is matched at most once. By the integrality of matching, w can be expressed as a convex combination of integral matchings. We randomly choose a matching according to the distribution defined by the combination. Let $C' \subseteq C$ be the set of matched clients. So, Property (4.7a) holds. For a fixed $S \in *$, the probability that $j \in C'$ is exactly $\sum_{i \in S} w_{i,j}$.

We now prove that Properties (4.7b) and (4.7d) hold simultaneously with probability at least $1/3$, conditioned on $S \in *$. To bound the probability for Property (4.7b), we focus on any $j \in C$ and bound its probability in C' .

$$\begin{aligned} \Pr[S \in *, j \in C'] &= \sum_{S \in *} z_S^B \sum_{i \in S} z_{S,i,j}^B / z_S^B = \sum_{S \in *, i \in S} z_{S,i,j}^B = x_{B,j} - \sum_{S \in \tilde{S} \setminus *, i \in S} z_{S,i,j}^B \\ &\geq x_{B,j} - \sum_{S \in \tilde{S} \setminus *} z_S^B = x_{B,j} - 1 + q, \end{aligned} \quad (15)$$

where the third equality is by summing up Constraint (8) for all facilities $i \in B$ and the inequality is by Constraint (11).

Let $C'' := \{j \in C : x_{B,j} \geq 1 - q\}$. For any $j \in C''$, we have

$$\Pr[j \notin C' \mid S \in *] = 1 - \frac{\Pr[S \in *, j \in C']}{\Pr[S \in *]} \leq 1 - \frac{x_{B,j} - 1 + q}{q} = \frac{1 - x_{B,j}}{q}.$$

$$\begin{aligned} \text{Thus, } \mathbb{E}[x_{B,C \setminus C'} \mid S \in *] &= \sum_{j \in C} x_{B,j} \Pr[j \notin C' \mid S \in *] \leq \sum_{j \in C''} x_{B,j} \frac{1 - x_{B,j}}{q} + \sum_{j \in C \setminus C''} x_{B,j} \\ &\leq \frac{1}{q} \sum_{j \in C} x_{B,j} (1 - x_{B,j}) = \frac{\pi(J)}{q}, \end{aligned} \quad (16)$$

where the equality is by linearity of expectation, and the second inequality is by $1 - x_{B,j} > q, \forall j \in C \setminus C''$.

Now, focus on Property (4.7d). For every $i \in B, j \in C$, we use $j \rightarrow i$ to denote the event that j is pre-assigned to i . Then $\Pr[S \in *, j \rightarrow i] = \sum_{S \in *} z_S^B z_{S,i,j}^B / z_S^B = \sum_{S \in *} z_{S,i,j}^B \leq x_{i,j}$, where the inequality is by Constraint (8). Thus, $\Pr[j \rightarrow i \mid S \in *] \leq x_{i,j}/q$. Therefore,

$$\mathbb{E}[\text{total connection cost} \mid S \in *] = \sum_{i \in B, j \in C} d(i, j) \Pr[j \rightarrow i \mid S \in *] \leq \frac{1}{q} \sum_{i \in B, j \in C} d(i, j) x_{i,j} = D_B/q.$$

We set $\ell_2 = \Theta(\ell \log \ell)$ to be large enough so that $3/q \leq \ell_2$. By Markov inequality,

$$\Pr[\text{total connection cost} \geq \ell_2 D_B \mid S \in *] \leq 1/3.$$

Also, by Markov inequality and Inequality (16), we have

$$\Pr[x_{B,C \setminus C'} \geq 3\pi(J)/q \mid S \in *] \leq 1/3.$$

By the union bound, with probability at least $1/3$, Property (4.7d) holds and $x_{B,C \setminus C'} \leq 3\pi(J)/q \leq \ell_2 \pi(J)$, which is exactly Property (4.7b). Since J is concentrated, this implies $x_{B,C \setminus C'} \leq 3x_{B,C}/(\ell_2 q)$. Notice that $Y - |S| \geq f(t)/2 \geq \frac{y_B}{2\ell q \delta}$. We set $\ell_2 = \Theta(\ell \log \ell)$ to be large enough so that $6\ell \delta \leq \ell_2$. Then,

$Y - |S| \geq \frac{3y_B}{\ell_{2q}}$, implying that $\frac{x_{B,C \setminus C'}}{x_{B,C}} y_B + |S| \leq \frac{3y_B}{\ell_{2q}} + |S| \leq Y$. This is exactly Property (4.7c). \square

A.6 Proof of Claim 4.8

Proof. Clearly, $\tilde{\alpha}_i \in [0, u_i]$ for every $i \in U_{P \setminus r}$. For any node $p \in P \setminus r$, we have $\tilde{\alpha}(U_p) = \frac{x_{U_p, \tilde{C}}}{x_{U_p, C}} x_{U_p, C} = x_{U_p, \tilde{C}}$. Summing up the inequalities over all $p \in P \setminus r$ gives us $\tilde{\alpha}(U_{P \setminus r}) = x_{U_{P \setminus r}, \tilde{C}}$.

Focus on a node $p \in P \setminus r$. We have $\sum_{i \in U_p} \frac{\tilde{\alpha}_i}{u_i} = \frac{x_{U_p, \tilde{C}}}{x_{U_p, C}} \sum_{i \in U_p} \frac{x_{i, C}}{u_i} \leq \frac{x_{U_p, \tilde{C}}}{x_{U_p, C}} \sum_{i \in U_p} y_i = \frac{x_{U_p, \tilde{C}}}{x_{U_p, C}} y(U_p)$. If J_p is not concentrated, the bound is at most $y(U_p)$. If J_p is concentrated, then let t_p be the number of pre-opened facilities in U_p using Lemma 4.7. Then, $\sum_{i \in U_p} \frac{\tilde{\alpha}_i}{u_i} + t_p \leq \frac{x_{U_p, \tilde{C}}}{x_{U_p, C}} y(U_p) + t_p \leq (1 + \frac{1}{\ell}) y(U_p)$. Summing up the bounds over all $p \in P \setminus r$, we have $\sum_{i \in U_{P \setminus r}} \frac{\tilde{\alpha}_i}{u_i} + t \leq (1 + \frac{1}{\ell}) y(U_{P \setminus r})$. So, all the constraints of LP(14) are satisfied. \square